

Description

SYSTEM AND METHOD FOR INTERNET BROADCASTING OF MPEG-4-BASED STEREOSCOPIC VIDEO

Technical Field

- [1] The present invention relates to a Web broadcasting system and method; and, more particularly, to a system and method for broadcasting a stereoscopic video to users on the Internet based on Moving Picture Experts Group (MPEG)-4.

Background Art

- [2] 'A Stereoscopic video' means a moving picture that is produced by receiving and outputting left-eye data and right-eye data alternately to give three-dimensional far and near distance effect to two-dimensional planes.
- [3] Along with the recent development of the Internet, diverse multimedia data in a field of education, culture, current issues and the like are provided to Internet users. Internet users can watch and/or listen to multimedia data they want at any time at any place as long as they have clients connected to the Internet.
- [4] Generally, Internet broadcasting systems, which are also referred to as Internet broadcasting systems, are formed of an encoding server for encoding multimedia data based on a predetermined encoding method, a streaming server for transmitting the multimedia stream, and clients for decoding and outputting the transmitted multimedia stream .
- [5] Fig. 1 is a block diagram illustrating a typical Internet broadcasting system. As shown, video data and audio data are inputted from a video/audio input device 10, such as a video camera, and compressed as they pass through an encoding server 20.
- [6] The MPEG is a group of moving picture experts that is formed to establish the standards for moving picture encoding methods. The MPEG studies about moving picture compression that varies continuously based on time and about the transmission of coded data. The MPEG suggests international encoding standards and current Internet broadcasting is performed based on the standards. Particularly, MPEG-1 and MPEG 2 are international standards that are used for compressing and storing large volume multimedia data.
- [7] A streaming server 30 transmits the multimedia stream, which is encoded by the encoding server 20, to clients 50 through the Internet 40. Then, the clients 50 decode the transmitted multimedia stream. The clients 50 should have a player with a codec to

output the multimedia data.

- [8] In the meantime, some problems may occur when the stereoscopic video data are transmitted using conventional encoding methods and current Internet broadcasting system. Since left-eye images and right-eye images should be encoded separately to transmit stereoscopic video data to the clients through the Internet, the amount of data is increased more than twice and the probability of transmission error becomes higher due to the load of transmission traffic. Moreover, there is a problem that the clients should discriminate between the left-eye images and the right-eye images in order to decode them and output them synchronized with each other temporally. If the left-eye images and the right-eye images are not outputted alternately, a three-dimensional effect cannot be obtained, only to cause eye-fatigue of viewers.

- [9] Therefore, a new encoding method, other than conventional encoding methods, is required to broadcast stereoscopic video data on the Internet as well as an Internet broadcasting system and method coinciding with the encoding method.

Disclosure of Invention

Technical Solution

- [10] It is, therefore, an object of the present invention to provide a system and method for broadcasting stereoscopic video data on the Internet by encoding and multiplexing multimedia data based on a structure of Moving Picture Experts Group-4 (MPEG-4) temporal scalability (TS).
- [11] It is another object of the present invention to provide an Internet broadcasting system and method that can broadcast conventional two-dimensional video data on the Internet.
- [12] In accordance with one aspect of this invention, there is provided a system for broadcasting stereoscopic video data to a client on the Internet, including: an encoding server for encoding stereoscopic video data, audio data, and Object Descriptor/Binary Format for Scene (OD/BIFS), which is information for controlling a content, and encoding the data into elementary stream (ES) having an MPEG-4 structure; a web server for receiving from the client any one among two-dimensional video display mode, field-shuttering video display mode and frame-shuttering video display mode; and a streaming server for generating a real-time transport protocol (RTP) packet for real-time data transmission on the Internet by multiplexing the ES based on the display mode inputted into the web server, and transmitting the RTP packet to the client.
- [13] In accordance with one aspect of the present invention, there is provided a method for broadcasting stereoscopic video data to a client on the Internet based on MPEG-4,

including the steps of: a) receiving stereoscopic video data, audio data, and OD/BIFS data, which is information for controlling a content, and encoding the data into ES having an MPEG-4 structure; b) receiving any one among two-dimensional video display mode, field-shuttering video display mode and frame-shuttering video display mode from the client; and c) generating an RTP packet for real-time transmission on the Internet by multiplexing the ES based on the inputted display mode, and transmitting the RTP packet to the client.

Description of Drawings

- [14] The above and other objects and features of the present invention will become apparent from the following description of the preferred embodiments given in conjunction with the accompanying drawings, in which:
- [15] Fig. 1 is a block diagram illustrating a typical Internet broadcasting system;
- [16] Fig. 2 is a block diagram depicting an Internet broadcasting system in accordance with a preferred embodiment of the present invention;
- [17] Fig. 3 is a block diagram showing an encoding server of Fig. 2 in detail;
- [18] Fig. 4 is a block diagram showing an encoder of Fig. 3 in detail;
- [19] Fig. 5 is a diagram showing a video data inputted into each layer of a Moving Picture Experts Group-4 (MPEG-4) structure in accordance with the preferred embodiment of the present invention;
- [20] Fig. 6 is a block diagram illustrating an MPEG 4 (MP4) file generator of Fig. 3 in detail;
- [21] Figs. 7 and 8 are diagrams describing arrangements of elementary stream (ES) of an MP4 file;
- [22] Fig. 9 is a block diagram illustrating a streaming server of Fig. 2 in detail; and
- [23] Fig. 10 is a diagram depicting a packing transformation process in the streaming server.

Mode for Invention

- [24] Other objects and aspects of the invention will become apparent from the following description of the embodiments with reference to the accompanying drawings, which is set forth hereinafter. The terms and words used in the present specification and claims should not be construed as conventional or dictionary meaning, but they should be construed as concepts and meanings fit in with the technological concept of the present invention based on a principle that inventors could define the concept of terms properly to describe the invention most appropriately. Accordingly, the embodiment and drawings of the present specification are no more than one of the preferred em-

bodiments and do not represent all the technological concept of the present invention. In the respect, there may be various equivalents and modifications that can replace the elements illustrated in the specification as of the filing of the present patent application.

[25] Fig. 2 is a block diagram depicting an Internet broadcasting system in accordance with a preferred embodiment of the present invention. As shown, multimedia data (i.e., stereoscopic video data or audio data) or content-controlling Object Descriptor/Binary Format for Scenes (OD/BIFS) data obtained from a stereoscopic video camera or a video/audio input device 100, such as video tape recorder (VTR), are inputted into an encoding server 200. Then, the encoding server 200 encodes the inputted signals based on Moving Picture Experts Group-4 (MPEG-4). An elementary stream (ES) obtained by encoding the signals in the encoding server 200 is transmitted to a streaming server 300.

[26] To encode the stereoscopic video, the present invention uses an MPEG-4 temporal scalability (TS). MPEG-4 TS is a structure where inputted left-eye images are allocated to a base layer and right-eye images are allocated to the enhancement layer. The left-eye images allocated to the base layer are encoded based on the conventional two-dimensional video encoding. The right-eye images allocated to the enhancement layer are encoded with reference to the image of the base layer, which is overlapped with that of the environment layer.

[27] Meanwhile, a web server 400 receives information on contents and a display mode requested by a client 600 through a back channel and transmits them to the streaming server 300. The streaming server 300 multiplexes the ES of the content in the display mode requested by the client 400 to generate multimedia data, e.g., a real-time transport protocol (RTP) packet, and transmits the multimedia data to the client 600 through the Internet. The client 600 decodes and displays the data in the transmitted order. To output multimedia data, the client 600 should have a player with a codec necessarily.

[28] Fig. 3 is a block diagram showing an encoding server of Fig. 2. As shown, the encoding server 200 includes an encoder 210, an encoding parameter unit 220, an MPEG layer 4 (MP4) file generator 230 for generating an MP4 file by using encoded ES, and a storage 240 for storing the MP4 file.

[29] The encoding parameter unit 220 provides information for encoding the inputted stereoscopic video. It sets up parameters for encoding, such as a size of an image, the number of frames to be encoded, a frame rate, a size of motion search, a transmission

bit rate, and an initial quantization coefficient, and inputs them to the encoder 210.

[30] The encoder 210 encodes the inputted stereoscopic video data and audio data based on the MPEG-4 TS and audio codec. Internal modules of the encoder 210 are illustrated in Fig. 4.

[31] Referring to Fig. 4, the encoder 210 includes a video encoding module 212 for encoding stereoscopic video data, an Elementary Stream Interface (ESI) information generating module 216, an audio encoding module 218 for encoding audio data, and an OD/BIFS encoding module 219 for encoding OD/BIFS data.

[32] The OD/BIFS encoding module 219 encodes binary format for scene (BIFS) for describing audio and scenes and object descriptor (OD) for defining the relationship between media streams.

[33] The ESI information generating module 216 generates additional information needed for the transmission and decoding of ES, such as a data length of ES, an idle flag, and a length of access unit (AU), which are included in a header information of a synchronization layer (SL). The header information of SL will be described later.

[34] The video encoding module 212 further includes a field separating module 213, a base layer encoding module 214, and an enhancement layer encoding module 215. The field separating module 213 separates a stereoscopic three-dimensional video data into a left-eye odd field, a left-eye even field, a right-eye odd field, and a right-eye even field. The base layer encoding module 214 encodes the left-eye odd field, and the enhancement layer encoding module 215 encodes the left-eye even field, right-eye odd field and right-eye even field.

[35] Fig. 5 is a diagram showing fields separated by the field separating module being inputted into each layer of an MPEG-4 structure in accordance with the preferred embodiment of the present invention. As shown, the left-eye odd field is inputted into the base layer; the left-eye even field into a first enhancement layer; the right-eye odd field into a second enhancement layer; and the right-eye even field into a third enhancement layer.

[36] Fig. 6 is a block diagram illustrating an MP4 file generator of Fig. 3. As shown, the MP4 file generator 230 which receives video/audio ES, OD/BIFS ES and ESI information from the encoder 210 includes a media data providing module 232, a metadata providing module 234 and an MP4 file generating module 236.

[37] The media data providing module 232 is a buffer for receiving video ES, audio ES and OD/BIFS ES, which are encoded on a field-by-field basis. It transmits the ES as to the MP4 file generating module 236.

- [38] The metadata providing module 234 is a buffer for receiving ESI information transmitted from the encoder 210, and transmitting the ESI information as a metadata to the MP4 file generating module 236.
- [39] The MP4 file generating module 236 converts the inputted ES and the metadata into an MP4 file format. This is to generate and store a file of a format suitably for transmission by receiving ES outputted from the encoder and additional information for the ES, extracting ES in coincidence with the display mode requested by a user.
- [40] An MP4 file has two zones: One is a metadata zone for storing file information, and the other is an mdata Atom zone for storing ES. The ES stored in the mdata Atom zone is given a proper ED identification ES_ID to discriminate encoded ES.
- [41] Figs. 7 is an exemplary diagram illustrating an arrangement of ES in the mdata Atom for storing the media data, the ES being given four ES_ID based on the right and left odd and even fields. Fig. 8 is an exemplary diagram illustrating an arrangement of ES for stereoscopic video data in the mdata Atom by multiplexing four fields of the ES. The ES is inputted on a four-field basis, i.e., a left-eye odd field, a right-eye even field, a left-eye even field and a right-eye odd field. One ES_ID is allocated to four fields having the same time information.
- [42] The MP4 file generated through the above processes is stored in a storage 240 and extracted by the streaming server 300.
- [43] Fig. 9 is a block diagram illustrating a streaming server of Fig. 2. As shown, the streaming server 300 extracts MP4 files stored in the storage 240, or receives encoded ES and ESI information encoded by the encoder 210, generates a real-time transport protocol (RTP) packet that coincides with a user's request, and transmits it to a client 600.
- [44] In order to generate the RTP packet that coincides with the user's request, a display mode requested by the user should be inputted into the streaming server 300. Accordingly, the display mode requested by the user should be inputted from the client 600 and a web server 400 and then transmitted to a streaming server 300.
- [45] In the Internet broadcasting system of the present invention, a video data is encoded after divided into a left-eye odd field, a left-eye even field, a right-eye odd field and a right-eye even field. Therefore, conventional two-dimensional video data, field-shuttering three-dimensional video data or frame-shuttering three-dimensional video data can be all processed in this system.
- [46] For example, if a user wants the conventional two-dimensional video display, the streaming server 300 transmits a stream of the left-eye odd field and the left-eye even

field. If the user wants field-shuttering three-dimensional video display, it extracts and transmits a stream of the left-eye odd field and the right-eye even field. Likewise, if the user wants frame-shuttering three-dimensional display, it transmits a stream of all of the four fields.

- [47] If the user's request on the display mode is inputted into a MP4 file analyzing module 310 through the web server 400, the MP4 file analyzing module 310 extracts a needed AU stream and ESI information from the MP4 files stored in the storage 240. Here, the MP4 file analyzing module 310 can receive the AU stream and the ESI information from the encoder 210 in real-time.
- [48] When the MP4 file analyzing module 310 extracts the AU stream and the ESI information based on the request of the user, a SL packet generating module 320 generates an SL packet having a header and a payload for the extracted AU stream. The header of the SL packet is synchronization information for each packet and it is used to check continuity when data loss occurs. The header includes information for controlling time synchronization, such as time stamp. The payload of the SL packet is valid information that comes after the header. The payload includes the AU stream extracted by the MP4 file analyzing module 310.
- [49] The generated SL packet is inputted into a FlexMux packet generating module 330 FlexMux, and the FlexMux packet generating module 330 generates a FlexMux packet by adding a header that defines a packet type to the SL packet. The packet type means information for distinguishing video data from audio data.
- [50] The generated FlexMux packet is inputted into an RTP packet generating module 340. Then, the RTP packet generating module 340 generates an RTP packet that could be transmitted through the Internet in real-time.
- [51] The RTP packet is a protocol packet of a transport layer that makes it possible to transmit data on the Internet in real-time. The RTP packet can be generated by adding a header including information for real-time data transmission to a FlexMux packet.
- [52] Fig. 10 is a diagram depicting a packing transformation process in the streaming server. The RTP packet generated in the above is transmitted to a client 600 through the Internet in real-time, and a player mounted on the client 600 decodes the RTP packet and displays it.
- [53] If the packet is a field-shuttering three-dimensional video RTP packet, the player can produce three-dimensional distance effect by outputting a stream of left-eye odd field and a stream of right-eye even field in the transmitted order, instead of discriminating between left-eye odd field stream and right-eye even field stream and syn-

chronizing their time with each other and output them. In short, since the RTP packet multiplexed by the streaming server 300 is packetized in the order of necessary field streams based on the display mode requested by the user, the client 600 can output stereoscopic video data without an additional data processing.

[54] The Internet broadcasting system and method of the present invention can reduce the amount of data considerably by encoding a stereoscopic video data effectively, thus reducing the probability of transmission error occurrence. Therefore, it is possible to broadcast stereoscopic videos on the Internet in real-time.

[55] In addition, the Internet broadcasting system of the present invention can restore not only stereoscopic videos but also conventional two-dimensional videos based on the display mode requested by the user.

[56] While the present invention has been described with respect to certain preferred embodiments, it will be apparent to those skilled in the art that various changes and modifications may be made without departing from the scope of the invention as defined in the following claims.